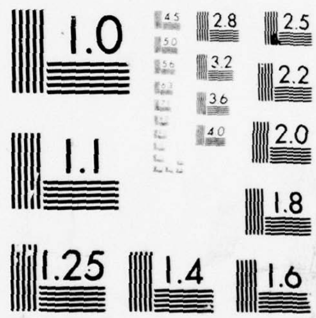
AD-A058 273 CALIFORNIA UNIV LOS ANGELES WESTERN MANAGEMENT SCIE--ETC F/G 5/1
MAKING BETTER USE OF OPTIMIZATION CAPABILITY IN DISTRIBUTION SY--ETC(U)
JAN 78 A M GEOFFRION N00014-75-C-0570
UNCLASSIFIED WMSI-WORKING PAPER-279 NL

1 OF 1
AD
A058273

END
DATE
FILMED
10-78
DDC



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

Working Paper No. 279

① LEVEL II

MAKING BETTER USE OF OPTIMIZATION CAPABILITY IN DISTRIBUTION SYSTEM PLANNING

by

ARTHUR M. GEOFFRION

January 1978

WESTERN MANAGEMENT SCIENCE INSTITUTE
University of California, Los Angeles

WESTERN MANAGEMENT SCIENCE INSTITUTE

University of California, Los Angeles

(14) WMSI- Working Paper [redacted]-279

(6)

MAKING BETTER USE OF OPTIMIZATION CAPABILITY IN DISTRIBUTION SYSTEM PLANNING

by

(10) ARTHUR M. GEOFFRION

(11) Jan[redacted]978

(12) 46p.

(15) N00014-75-C-0570

The author is pleased to acknowledge the constructive criticisms of Donald Erlenkotter, Glenn Graves, and Douglas Linton. The reproduction of this working paper was supported by the National Science Foundation and the Office of Naval Research. Portions were presented as part of a plenary lecture at the ORSA/TIMS/AIIE Distribution Conference, Hilton Head, South Carolina, February 1978.

072 320

46 LB

Abstract

To have an efficient optimization technique for a class of problems is to have no more than a tool. Like any tool, it can be used well or poorly. This paper is about how to use one such tool for distribution planning problems (see the companion piece by A. Geoffrion, G. Graves and L. Lee, "Strategic Distribution System Planning: A Status Report," Working Paper 272A, March 1978). Discussion centers on four topics of importance in practical applications: the relationship between system cost and the number of distribution facilities, sensitivity analysis, robustness analysis, and implementation priority analysis. Each of these topics requires the use of optimization in ways that are sometimes less than obvious. Several illustrations are drawn from actual applications in the auto parts, consumer products, food, and mining industries.

Good models and powerful optimization software do not a successful project make. Among the additional necessary ingredients is a generous measure of imagination and cunning regarding ways to use the model and software in support of ultimate project objectives. We are thus led to ask what types of optimization runs should be done and how the results should be interpreted.

A general conceptual discussion of this issue in the context of distribution system planning has been given elsewhere.[1/] Our purpose here is to give a more detailed discussion of four particular types of optimization runs which users often either overlook or have conceptual difficulty with. They are:

A. System Cost versus Number analysis

B. Sensitivity analysis

C. Robustness analysis

D. Implementation priority analysis.

There are, of course, many other important types of optimization runs beyond these four. Liberal use is made of numerical examples drawn from actual applications in the auto parts, consumer products, food and mining industries.

The reader is presumed to have a general familiarity with distribution system planning models of the type described in Ref. 5.

[1/] See Ref. 3, and also Sec. 4.2 of Ref. 1 and pp. 27-29 of Ref. 2.

A. THE SYSTEM COST VS. NUMBER CURVE

The *System Cost vs. Number curve* (CvN for short) gives the minimum possible total system cost as a function of the number of open distribution centers. Figure 1 gives an example of such a curve for the packaged goods division of a mining company (the model has 23 product groups, 12 plants, 51 candidate distribution center locations, and 110 customer zones). The lowest system cost in this example is obtained with a certain system having $n^*=30$ facilities open. The cost implications of departing from this number of facilities can be seen at a glance. For instance, there are systems with as few as 22 and as many as 39 facilities that come within 1% costwise of the best 30 facility system.

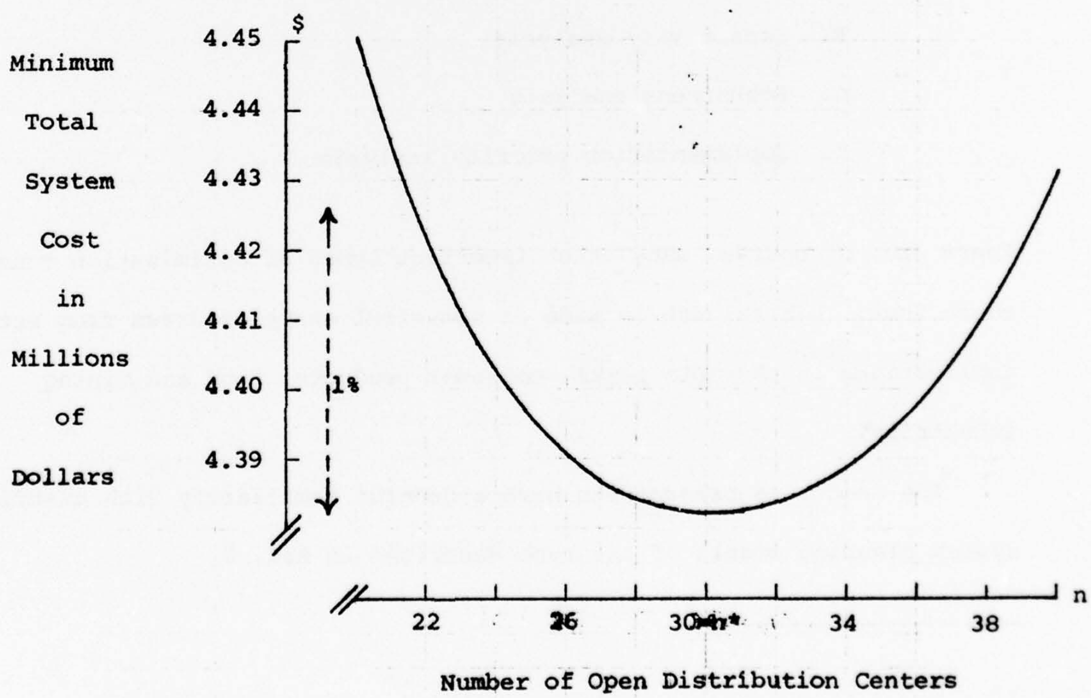


Figure 1

The System Cost Vs. Number (CvN) Curve For a Mining Company

Figure 1 and all similar figures are drawn for convenience as though n, the number of open facilities, were a continuous-valued rather than integer-valued variable.

It bears emphasis that this kind of curve cannot be obtained by the very common practice[2/] of simply costing out a number of plausible system configurations having different numbers of open facilities. That would just produce some points above the CvN curve. The ability to perform global cost minimization subject to a constraint on the number of facilities is essential in order to obtain a true CvN curve, which can be viewed as the lower envelope of a scatterplot of cost against size for all feasible system designs. Of course, all problem data must be held constant during the generation of this curve; different cost/demand/service scenarios lead to different curves.

There are at least three reasons why the CvN curve is valuable. These are taken up in turn.

1. First Rationale: The CvN Curve is of Direct Interest to Management

The number of open distribution centers is a distribution system design parameter of great intrinsic interest to the manager and system analyst. Whether a rebalanced system should have more or fewer facilities than the present system, and by how much the total number should change, is almost always a major issue. The CvN curve reveals how large a penalty is associated with departures up or down from the ideal number n* of facilities. Such departures can be desirable for a variety of hard-to-quantify reasons exterior to the formal model. To illustrate: (i) if the current number of facilities is substantially different from n*, management can make use of

2/ e.g., p. 227 of Ref. 6.

the penalty information to design a conservative initial strategy which goes only part way toward n*; (ii) management may prefer to have more than n* facilities for the sake of improved customer service; (iii) management may prefer to have fewer than n* facilities because the model has omitted certain economy-of-scale cost savings that were difficult to estimate.

2. Second Rationale: Getting at the Cost/Customer Service Tradeoff

The CvN curve facilitates making the difficult but important tradeoff between system cost and customer service level. This comes about because numerical generation of a CvN curve produces, as an inevitable by-product, some complete system designs that attain (or nearly attain) the curve at various points. The customer service characteristics of these system designs can be summarized in convenient terms, say the percentages of demand that can be filled within one-day, two-day, and three-day delivery. Unless there is some reason for considering systems with fewer than n* distribution centers, just the system designs with n* or more facilities need be so summarized since smaller systems have higher cost and generally worse customer service. Thus we obtain a series of alternative systems with progressively higher cost but with progressively better customer service. Management is then in a position to work out an informed compromise between the conflicting objectives of minimizing distribution cost and maximizing customer service.

The tradeoff possibilities between cost and customer service can be pictured graphically if the latter is measured by an index number such as the systemwide average (or median or maximum) distance between customers and their assigned distribution centers. For the firm whose CvN curve is shown in Figure 1, the demand-weighted average distance varies as shown in Figure 2. The cost/service tradeoff posibilities are shown most clearly if Figures 1 and 2 are combined as in Figure 3. This is the "cost-service tradeoff curve" which management must ponder.

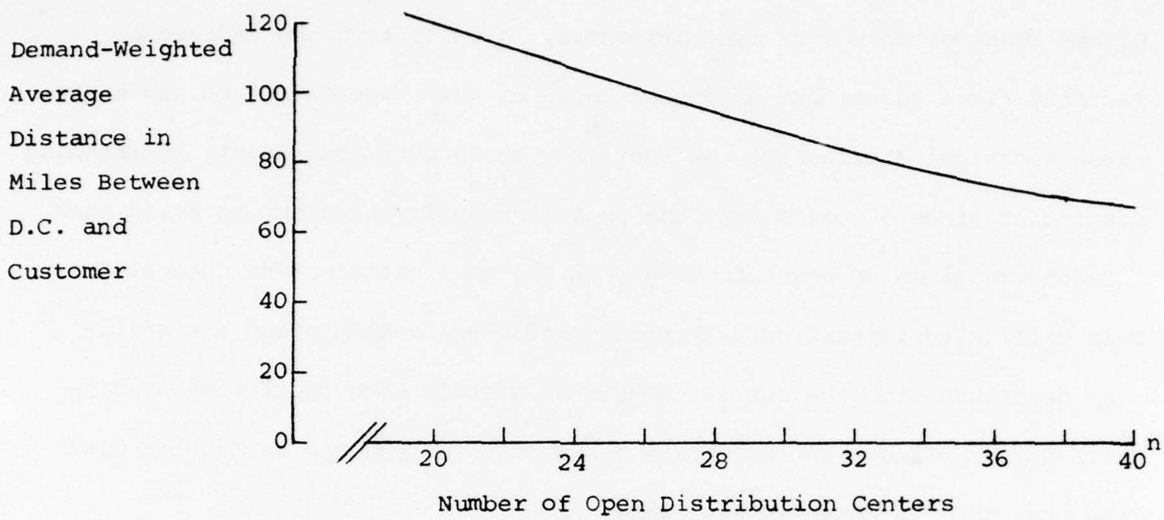


Figure 2

How A Proximity Measure of Customer Service Varies With The Number of Open Facilities (for a Mining Company)

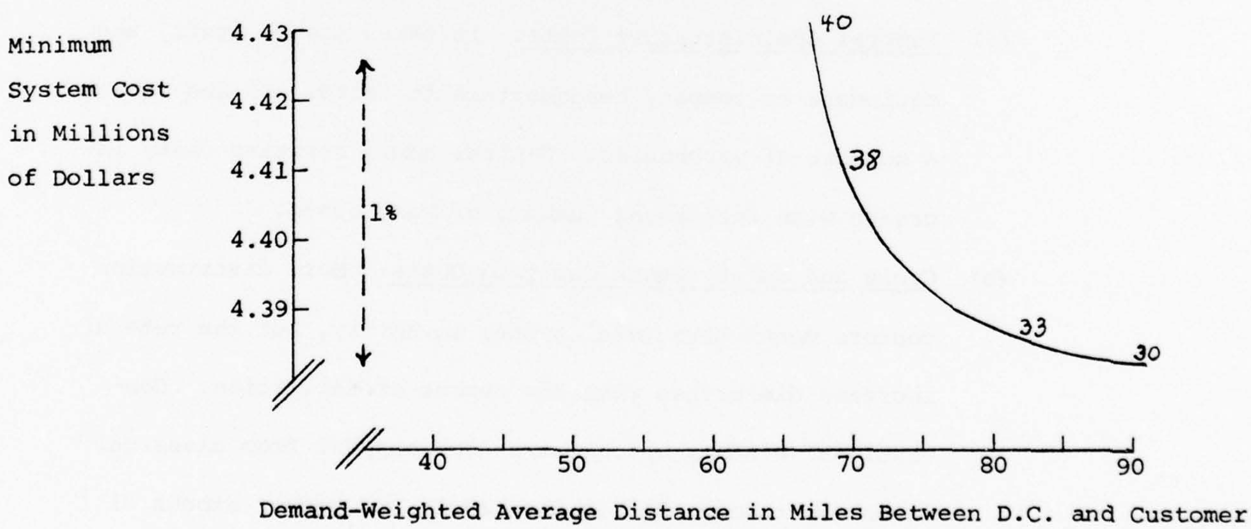


Figure 3

A Combination of Figures 1 and 2: How Minimum System Cost Varies With a Proximity Measure of Customer Service (for a Mining Company)

3. Third Rationale: A Device for Handling Problematical Costs and Savings That Are Dependent on the Number of Open Facilities

Usually there are significant categories of cost or potential savings which depend strongly on the number of open distribution centers. Linear costs of this sort can, of course, be built into the individual facility fixed costs, but nonlinear costs of this type cannot be and hence cause technical difficulty. Or there may be so much uncertainty surrounding particular kinds of costs that the project team is reluctant to build them into the model on an equal footing with the more certain cost categories. This difficulty is particularly pronounced when contemplating a significant departure from the current number of distribution facilities. As we shall see, nonlinear and uncertain costs of this type can be handled with ease once the CvN curve is available.

Here is a partial list of cost categories that can be modeled as functions of the number of open distribution centers:

(a) Central Administrative Costs. It takes space, staff, and equipment at company headquarters to coordinate and manage a network of warehouses. Central administrative costs increase with increasing numbers of warehouses.

(b) Cycle and Safety Stock Carrying Costs. More distribution centers means more total system inventory, but the rate of increase diminishes with the number of facilities. Conventional wisdom, which draws some support from classical inventory theory, has it that cycle and safety stocks will increase approximately in proportion to the number of facilities raised to an exponential power between 1/2 and 1. It is usually possible in principle to build inventory carrying costs directly into the distribution center throughput cost functions, but when numerous inventory assumptions or policies need to be tested (e.g., alternative assumptions on the cost of capital or policies regarding allowable stockout probability) it may be more expedient to model these costs as a function of the total number of open facilities.

(c) Consolidation Stock Carrying Costs. As the number of distribution centers increases and the average throughput decreases, there may be an increase in what might be called "consolidation" stocks. This refers to the need to accumulate a sufficient quantity of unfilled customer orders at a distribution center so that a reasonably economical delivery load can be assembled. The aim is to avoid the high costs of small shipments.

(d) Customer Order Size Effects. Customers which are close to a warehouse generally tend to order more frequently and in smaller quantities than customers which are farther away. This implies that delivery costs tend to increase on a $/cwt-mi basis as the number of facilities increases.

(e) Inter-warehouse Transfer Costs. The more distribution centers there are, the greater the coordination problems and hence the greater the tendency to transfer inventory between facilities in response to imbalanced inventory availability.

(f) Negotiated Reductions in Warehousing and Delivery Costs. The smaller the number of distribution centers, the greater their individual volume and hence the more opportunity there is to negotiate more favorable arrangements for warehousing and delivery services. Many modelers are reluctant to build such sliding cost reductions into the problem data because of uncertainty as to their magnitude.

Including such costs in the model tends to make it more complete and hence a more powerful tool for problem-solving, but at the expense of making it more difficult to solve (owing to nonlinearities) and of rendering it less credible (owing to cost/savings uncertainties). This presents a dilemma. We propose to deal with this dilemma by leaving such costs out of the model, solving it parametrically on the number of open facilities, and then factoring in the omitted costs manually. There is no loss of global optimality with such an approach so long as the omitted costs depend just on the number of open facilities, and the manual part of the procedure can be repeated easily with a variety of estimates for the uncertain costs.

Suppose that the total dollar amount of the problematical costs is estimated for different numbers of open facilities. Call these the n-costs for want of a better name, and make sure that they do not duplicate any cost components comprising system costs. The system design which minimizes the overall sum of system costs plus n-costs can be determined by simple arithmetic calculation once the CvN curve is in hand. Table 1 provides an illustration. Write the range of system sizes of interest in the first column, the corresponding CvN figures in the second, and the corresponding n-costs in the third. Add columns 2 and 3 to obtain column 4. Scan down column 4 to find the smallest overall cost. The corresponding system itself may already be known as a byproduct of having generated the CvN curve; otherwise (in the case where the key line is based on an interpolation of the CvN curve) the corresponding system can usually be found with little difficulty by the same technique used to generate the CvN curve.

Number of Open Facilities	System Costs from CvN Curve	Estimated n-Costs	Overall Costs	
24	$4,402,500	$ 80,000	$4,482,500	
26	4,390,500	86,000	4,476,500*	least overall cost
28	4,384,000	93,500	4,477,500	
30	4,382,000	100,000	4,482,000	
32	4,383,500	106,500	4,490,000	
34	4,388,500	113,500	4,502,000	
36	4,398,000	120,000	4,518,000	

Table 1

Sample Calculation of the Revised Optimal System Design When n-costs are Taken into Account (only every 2nd line is shown)

It should be obvious that nonlinearity of the n-costs offers no difficulty. Moreover, alternative assumptions for various n-cost components can be tried out easily without any additional computer runs (the CvN curve stays the same -- just the n-costs vary). Thus one may deal even with highly speculative costs in a convenient series of manual "what if..." trials. For instance, the inventory carrying cost component can be recalculated for several different stockout probabilities (say 10%, 5%, and 3%) and the impact on system design can thereby be assessed for this important measure of customer service.

The CvN curve is a remarkably powerful tool when used in this fashion.

4. Theoretical Curves

The general shape of a System Cost vs. Number curve can be derived analytically using calculus under highly simplified assumptions. Such an exercise is worthwhile for several reasons. First, the theoretical curve indicates sensible ways for interpolating between available points on the actual CvN curve and for extrapolating beyond them. Second, having the theoretical curve in advance facilitates the generation procedure described in the Appendix. But most importantly, a general agreement between the theoretical curve and the real one suggests why the CvN curve looks the way it does: the key cost tradeoffs, between decreasing outbound costs and increasing fixed costs as the number of open distribution centers increases, which underly the theoretical curve are probably similar for the full scale model.

It is a straightforward matter to derive a formula for the CvN curve for the simplified analytical model developed in the Appendix. The curve itself depends on the problem data in a fairly complicated way, but one can obtain a good idea of the general shape of the curve by (i) excluding those cost components which do not depend on n in the best n-facility system, and (ii) normalizing the curve to give relative system cost versus relative number of open facilities. The effect of (i) is to exclude variable facility costs and inbound transportation costs to the centroid of each open facility's service area. The remaining "non-excluded" costs can be thought of, loosely speaking, as fixed and outbound costs. "Relative" means that costs and the number of open facilities are expressed as a fraction of their optimal values. Under (i) and (ii), a single CvN curve covers all numerical cases[3/].

[3/] See the Appendix for a derivation. With the help of (3a)-(3d) in the Appendix and the curve shown in Figure 4, it is a simple matter to construct the full CvN curve with no cost components excluded.

The result is given in Figure 4, where $TC^*(n)$ denotes the total non-excluded costs for the best n-facility system. The optimal number of open facilities is denoted by n^*. Figure 4 shows the relative total cost $TC^*(n)/TC^*(n^*)$ as a function of n/n^*. Thus if one contemplates using the least cost system design having half the ideal number of facilities ($n/n^* = 0.5$), fixed plus outbound costs are predicted to be 11% higher than for the least cost system design having n^* facilities (since $TC^*(n^*/2)/TC^*(n^*) = 1.11$).

How closely does this curve approximate real CvN curves? Often fairly well. The little triangles in Figure 4 correspond to points on the real CvN curve for the mining company model mentioned earlier. The fit is quite decent. It would seem the counterplay between fixed and outbound costs in the simplified analytical model may bear a resemblance to the situation for the full mixed integer linear programming model, where the thousands of individual data elements tend to obscure what is really going on.

Not only can theoretical CvN curves be derived, but proximity-based customer service relationships like the one in Figure 2 can also be derived under simplifying assumptions. This is worthwhile for reasons similar to those mentioned earlier.

It is easy to show that in the case of equal hexagonal service areas covering uniformly distributed demand, the median (or average or maximum) distance between customers and their assigned distribution centers must

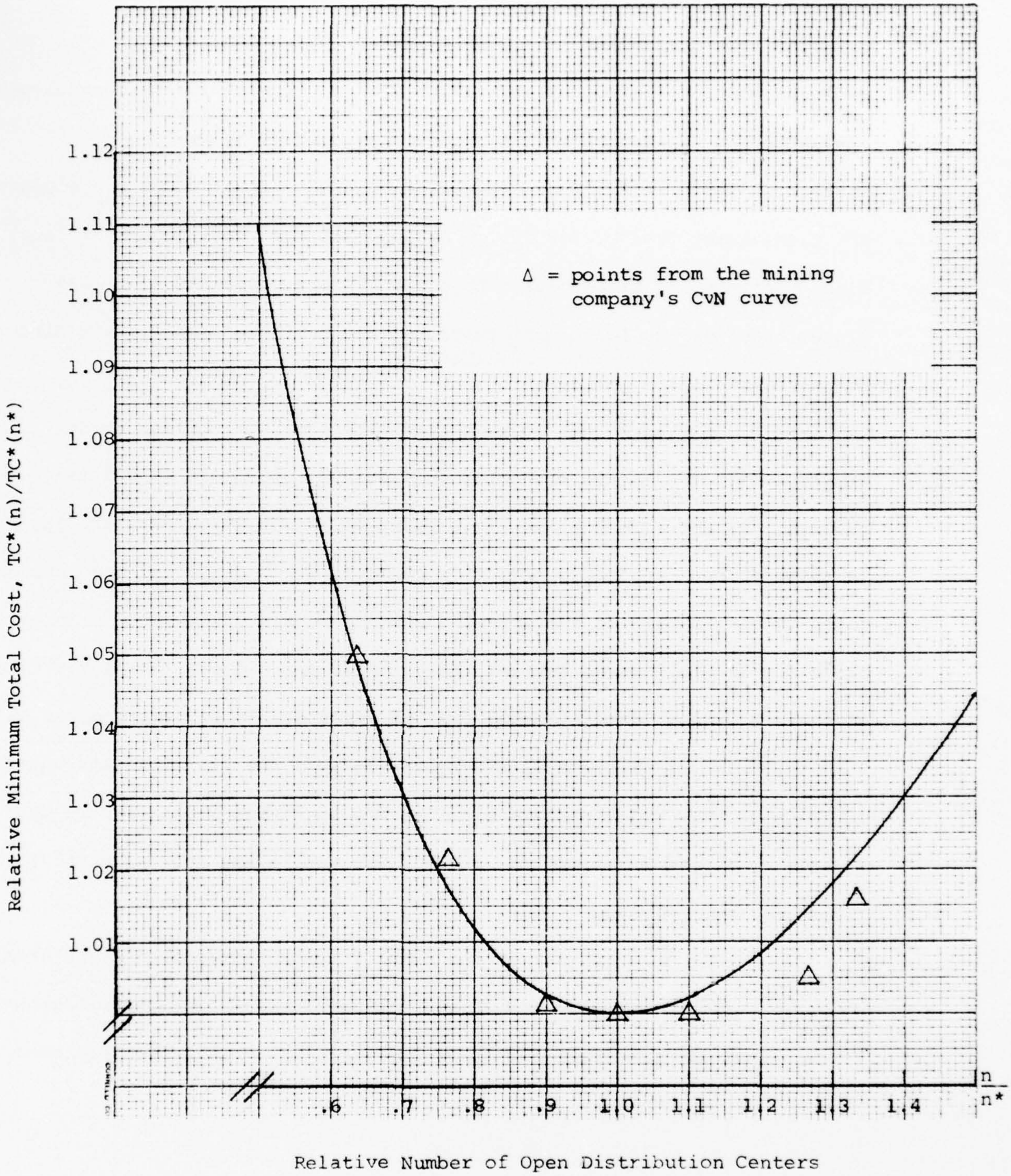


Figure 4

Normalized Theoretical CvN Curve
(derived in the Appendix)

vary as one over the square root of the number of facilities.[4/] Figure 5 shows this relationship for the continental U.S. (omitting 33% of the land area as too sparsely populated to require coverage). The striking feature of this curve is that the marginal advantage of additional distribution centers becomes trifling once there are more than 15 or 20 of them.

It is interesting to combine Figures 4 and 5 together in the same way Figures 1 and 2 were combined. This is done in Figure 6, which shows the theoretical tradeoff between a proximity measure of customer service and system cost (more precisely, the fixed plus outbound freight component of system cost). Several curves are drawn because the tradeoff depends on the optimal number n^* of facilities, which in turn depends on problem data. The numbers written adjacent to each curve indicate how many facilities correspond to selected points on the curve.

A conspicuous feature of these curves is that the cost penalty necessary to obtain a significant improvement in customer service quickly becomes prohibitive for values of n greater than n^* by more than about 20%.

[4/] Let n equal hexagons, each centered on a distribution center, cover an area of A square miles. The median distance from the center of each hexagon is that number r which satisfies $\pi r^2 = 0.5\ A/n$, that is, the area within r miles of the center of each hexagon equals half of the hexagon's total area. Thus $r = \sqrt{A/2\pi}\ /\sqrt{n}$.

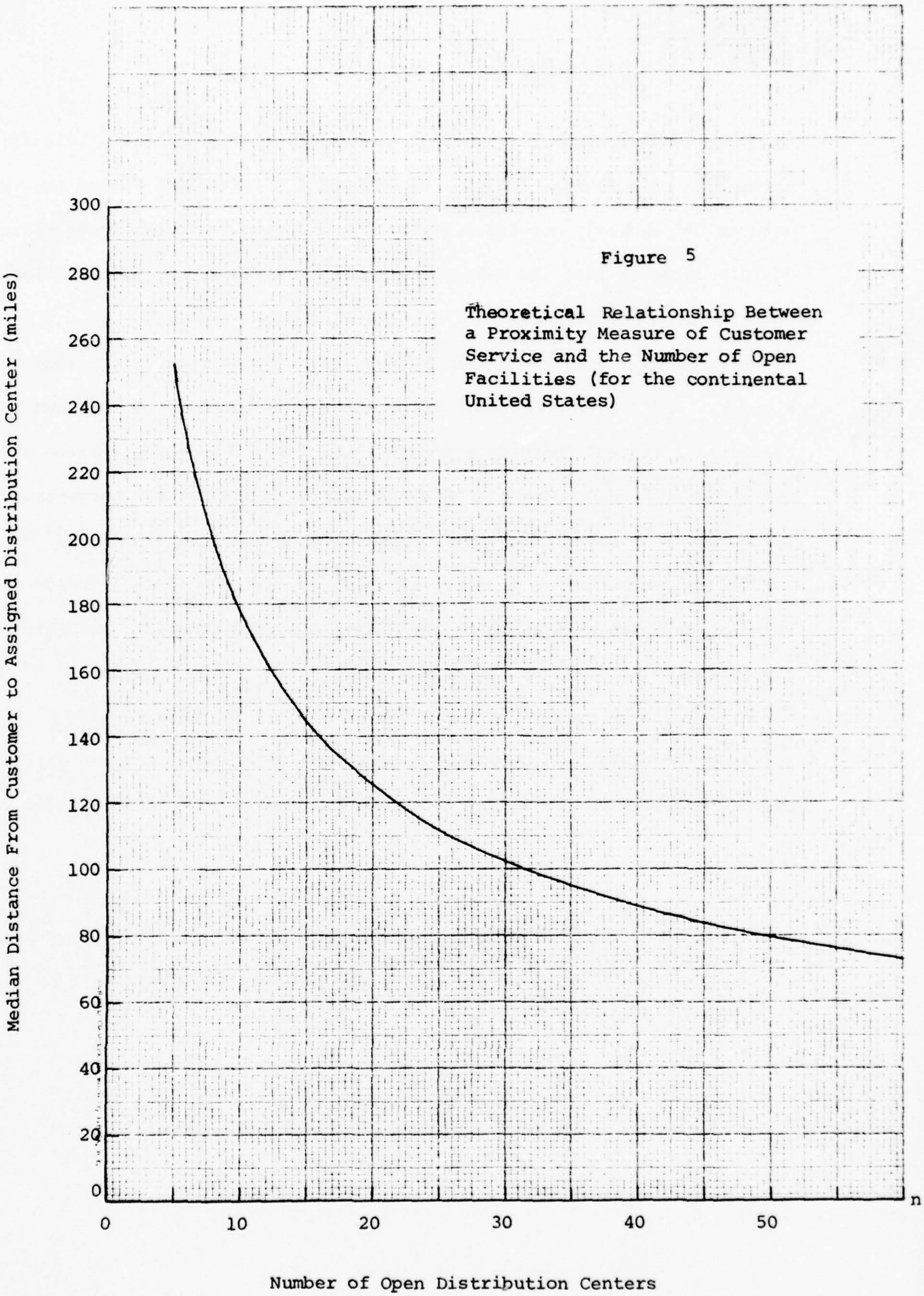


Figure 5

Theoretical Relationship Between a Proximity Measure of Customer Service and the Number of Open Facilities (for the continental United States)

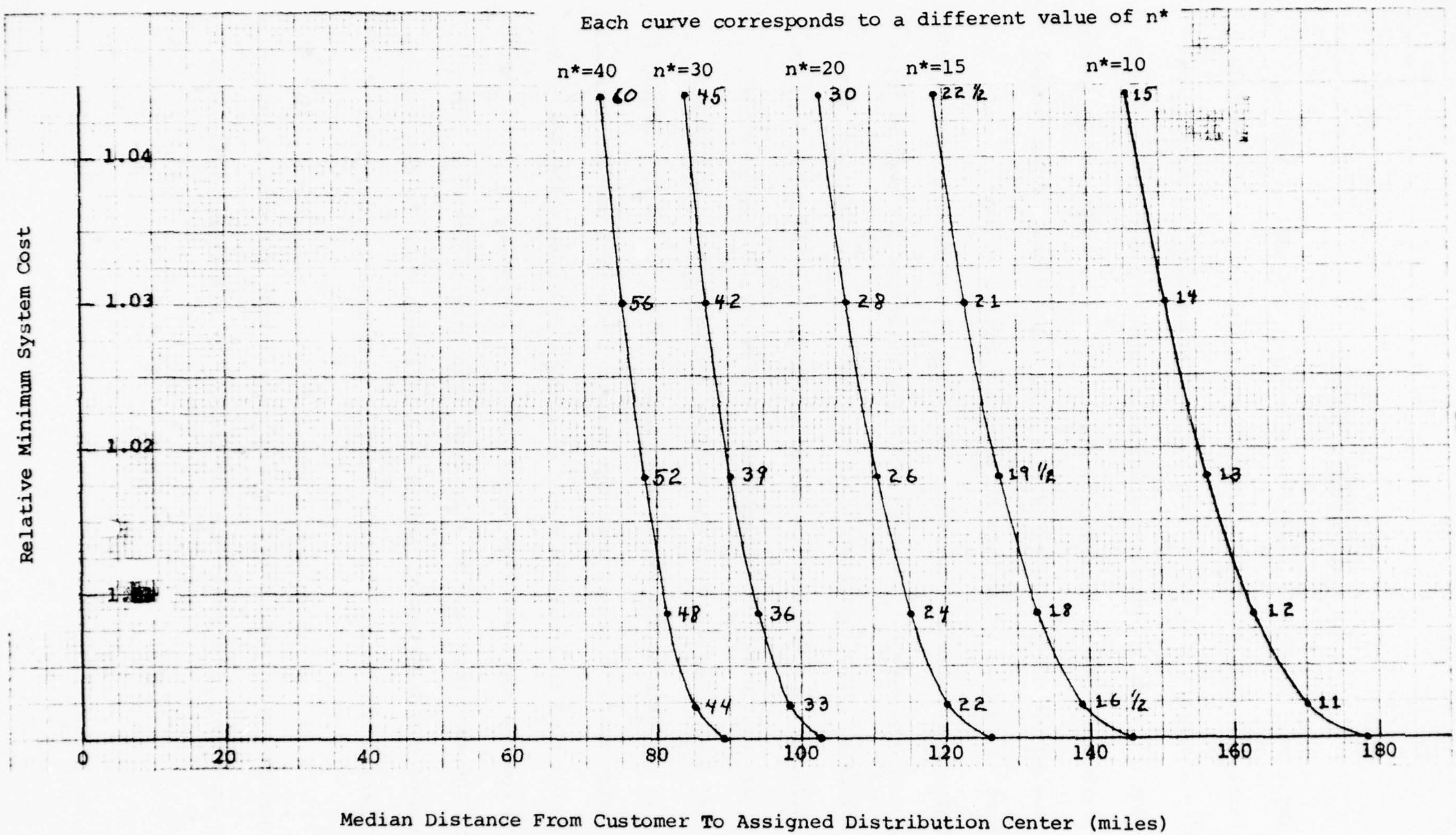


Figure 6

Combination of Figures 4 and 5: How Relative Minimum System Cost (Fixed + Outbound Freight) Varies With a Proximity Measure of Customer Service [annotations indicate the number of open distribution centers]

B. SENSITIVITY ANALYSIS

Empirical *sensitivity analysis* is based on the idea of setting up a family of runs for which selected components of the data are altered by a formula that depends on a single adjustable parameter. To carry out a series of sensitivity analysis runs, one optimizes for each of a series of settings of the parameter. Often the parameter is just a scale factor for certain model data, say outbound costs. Other times the parameter enters additively or in more complicated ways.

There are numerous situations where sensitivity analysis is in order. One is when the modeler is uncertain of the estimates for certain types of model data. Sensitivity analysis on these data may show that there is no significant dependence of the model solution over the range of uncertainty; or, in the event of significant dependence, knowledge of the extent of dependence may help the modeler decide how much additional effort to allocate to the estimation task. Another situation calling for sensitivity analysis occurs when the analyst feels that certain data will change over time in a certain way. It is usually straightforward to determine how the model solution would change in response. A third use, perhaps the most important of all, is to help develop insights into the workings of the distribution system -- a tool with which to find out the "whys" behind the "whats" in each stack of optimization output.

The variety of particular types of sensitivity analysis is potentially enormous. Discussion will be limited to systematic variations in:

- inbound freight rates
- outbound freight rates
- distribution center fixed costs
- distribution center variable costs
- demands

Each of these categories involves multiple data elements, sometimes more than a thousand. Usually one thinks in terms of parametric changes applied to selected populations of numbers rather than to individual numbers themselves. It therefore seems natural to take into account how any proposed parametric change will alter the two most fundamental descriptive measures of each population: its mean and its standard deviation. This distinction turns out to be particularly appropriate because the influence of one of these measures, the mean, allows a considerable amount of theoretical understanding whereas the other most definitely does not (at least at the present time).

Now ordinary multiplicative scaling, while of appealing simplicity and sometimes having a plausible rationale, has the unfortunate property that it factors both the mean and the standard deviation of the population so scaled. One may avoid this confounding effect by parameterizing in some other way which changes the mean or the standard deviation, but not both at the same time.

The proper type of parameterization for studying the influence of the mean (we'll call this the mean effect) is simple additive change. Just add a constant to all of the data elements to be changed. The mean is correspondingly changed but the standard deviation is not, nor is any other shape parameter of the distribution.

The proper type of parameterization for studying the influence of the standard deviation (we'll call this the dispersion effect) is not quite so obvious, but a good choice is summarized as follows.

Proposition Let the numbers $\{x_1, x_2, \ldots, x_n\}$ have mean μ and standard deviation σ . Define

$$y_i = \mu + \beta(x_i - \mu), \ i = 1, \ldots, n$$

where $\beta \geq 0$. Then the numbers $\{y_1, y_2, \ldots, y_n\}$ have mean μ and standard deviation $\beta\sigma$.

This simple parameterization factors the standard deviation but leaves the mean unchanged.

Empirical sensitivity results can be obtained for these two types of parameterization (or any other type) by brute force, that is, by optimizing for each of a series of parameter settings. A high degree of solution regularity is often observed when this is done, particularly if the mean effect is dominating. This suggests that idealized relationships might be derived to guide the interpolation and extrapolation of the empirical results. Since each point on a sensitivity curve requires a full optimization run, success along these lines could be of considerable practical utility. Moreover, the analytical derivations themselves of the theoretical sensitivity curves are likely to deepen our understanding of why the curves have the general shape that they do.

It is possible, under highly simplified assumptions, to derive an analytical expression for the optimal number of facilities. Such an expression permits sensitivity analysis predictions if one hypothesizes that the proportional change in the number of facilities in the full scale model will be the same as the proportional change in the number of facilities in the simplified analytic model.

A useful analytic minimodel for one plant and one product is explained in the Appendix, where it is demonstrated that the following formula is a good approximation to the optimal number of warehouses covering a total area of A mi^2 at minimum total cost:

$$n^* = .33\ A\ (\rho t/f)^{2/3}\ \left(1 - (r/t)^2\right)^{1/2}.$$

The symbols are defined as follows:

- ρ uniform demand density, in CWT/mi^2yr
- f fixed cost of an open warehouse, in \$/yr
- r inbound freight rate, in \$/CWT mi
- t outbound freight rate, in \$/CWT mi

For this formula it is easy to predict the <u>mean</u> effect of A, ρ, r, t, or f. For instance, changing f to $\tilde{f}$ changes n* by the factor

$$\left.\frac{\tilde{n}^*}{n^*}\right|_{\substack{\text{from the}\\ \text{minimodel}}} = \frac{.33\ A\ (\rho t/\tilde{f})^{2/3}\left(1 - (r/t)^2\right)^{1/2}}{.33\ A\ (\rho t/f)^{2/3}\left(1 - (r/t\ ^2\right)^{1/2}} = (f/\tilde{f})^{2/3}.$$

We therefore hypothesize

$$\left.\frac{\text{number of warehouses if } \tilde{f} \text{ is used}}{\text{number of warehouses if } f \text{ is used}}\right|_{\substack{\text{from the}\\ \text{full scale}\\ \text{model}}} \doteq (f/\tilde{f})^{2/3}$$

In a similar manner, we can develop the following hypothesis for the influence of the mean effect on inbound freight rates:

$$\left.\frac{\text{number of warehouses if } \theta r \text{ is used}}{\text{number of warehouses if } r \text{ is used}}\right|_{\substack{\text{from the}\\ \text{full scale}\\ \text{model}}} \doteq \left[\frac{1 - \left(\frac{\theta r}{t}\right)^2}{1 - \left(\frac{r}{t}\right)^2}\right]^{1/2}$$

These two formulas are graphed in Figure 7 along with the formulas for the influence of the mean effect for demand density and outbound freight rate. The "parameter change factor" is the ratio of new to old value for the parameter of interest; in the notation used above, for instance, the factor was $\tilde{f}/f$ in the case of fixed cost and θ in the case of inbound freight rate. The "n* change factor" is just the ratio of the new value of n* to the old value of n*. Thus the sensitivity curve for f indicates that 20% fewer warehouses are needed when fixed costs all increase by 40%.

Notice that there is just one sensitivity curve for f and ρ, but the curves for r and t depend on the ratio r/t. In practice this ratio is usually well under 1/2. Curves are given for ratio values of 0, .2, .3, .4 and .5. The ρ curve coincides with the t curve for r/t = 0.

The analyst should compare such predictions with the empirical results of sensitivity analysis. General agreement seems to occur more often than not, but disagreement can also be valuable because it forces an inquiry into the reasons for the discrepancy.

First Example

A series of runs for the mining company model was done in which different constants were added to each of the facility fixed costs. This is exactly the additive change parameterization recommended previously for studying the mean effect. In Table 2 we compare the actual change in n* with the theoretical prediction

$$\frac{n^*}{30} \cong \left(\frac{15{,}800}{\text{mean fixed cost}} \right)^{2/3} .$$

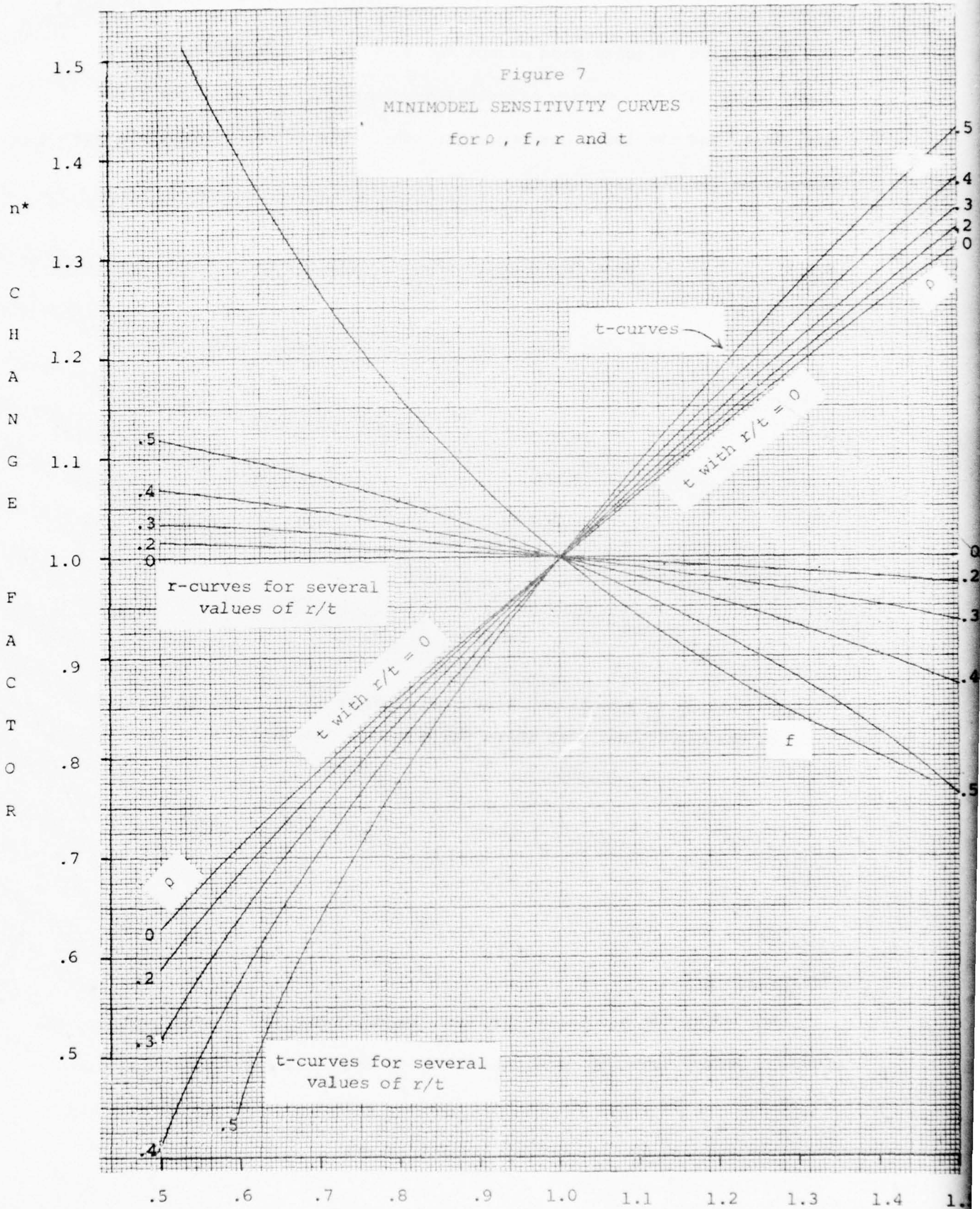


Figure 7
MINIMODEL SENSITIVITY CURVES
for ρ, f, r and t
n* CHANGE FACTOR
1.5
1.4
1.3
1.2
1.1
1.0
.9
.8
.7
.6
.5
t-curves
t with r/t = 0
r-curves for several values of r/t
t with r/t = 0
f
ρ
ρ
t-curves for several values of r/t
.5
.4
.3
.2
0
.5
.6
.7
.8
.9
1.0
1.1
1.2
1.3
1.4
PARAMETER CHANGE FACTOR

The prediction is quite good except for very large downward changes in fixed costs, where boundary effects (there are only 51 candidate sites available) and the lower warehouse throughput limits offer plausible explanations for the discrepancy.

Run #	Mean Fixed Cost	Actual n*	Predicted n*
1	$36,600	19	17.1
2	25,000	23	22.1
3	20,000	27	25.6
4	15,800	30	30.0
5	13,500	33	33.2
6	10,000	38	40.7
7	5,000	40	51

TABLE 2

Comparison of Actual and Predicted Changes in n* Owing to Changes in the Mean Fixed Cost (Data for a Mining Company)

Second Example

The model for a consumer products manufacturer has a single plant, one product group, about 65 candidate facility locations, and about 130 customer zones. An issue of interest was the sensitivity of the full scale model to changes in the inbound freight rates.

With an r/t ratio of 0.082, the analytic minimodel predicts a very low degree of sensitivity. It predicts, for instance, that a 50% reduction in inbound freight rates will increase the number of warehouses by only 1/4 of 1%.

This prediction was tested by subtractively reducing inbound freight rates by an average of 50% and then rerunning the full scale model. The result was that the optimal number of warehouses increased from 24 to 25, which is in quite reasonable agreement with the prediction.

It warrants emphasis that both the prediction and the empirical test pertain to the mean effect and <u>not</u> to the dispersion effect. We made subtractive rather than multiplicative changes when setting up the new run of the full scale model, but simply halving all inbound freight rates -- which, as we know from earlier discussion, involves both mean and dispersion effects -- increases the optimal number of warehouses by 4 to 28. Thus the dispersion effect has an influence not at all well predicted by the minimodel.

This illustration was adapted from Ref. 4, wherein a somewhat more elaborate extension of the minimodel is suggested which does appear to give useful predictions of dispersion effects as well as mean effects.

In closing, it bears emphasis that minimodels should be thought of as adjuncts to, and not as replacements for, full scale models. The most one can hope for is that they capture the essential features of the problem well enough to exhibit the general character of the solution in a qualitatively correct way and to reveal the basic forces responsible for the character of the solution. An effective minimodel can enhance the usefulness of a full scale model in much the same way that a small sighting scope enhances the usefulness of a large telescope.

C. ROBUSTNESS ANALYSIS AND REGRET

Before recommending a reconfigured distribution system it is wise to study its robustness in the face of various departures from the assumptions on which it is based. How will the proposed system fare five years from now when the demand pattern has changed? When a backhaul revenue development program reaches maturity? How would the proposed system fare if outbound freight rates increase twice as fast as inbound freight rates? If that perennial possibility of acquiring a new plant on the West Coast should materialize?

Such questions, and many others like them, challenge the validity of any proposed redesign of a distribution system. Sometimes they can be disposed of by logical argument or manual analysis without the need to perform additional optimization runs. But often the only way to address them properly is by some type of robustness analysis performed on the full scale model.

The typical robustness question is of the following type:

> *"Proposed system A is best for model data scenario X; would system A still be nearly best if some other data scenario Y should materialize?"*

A typical naive approach to this question is to optimize using data scenario Y to find the best system B in that case. Systems A and B are then compared; if the differences look fairly "small", then system A is declared to be "robust" with respect to Y. Another common naive approach is to price out (no optimization) system A using scenario Y; if the cost doesn't change much, system A is declared to be robust.

Both of these approaches can be misleading. The trouble with them is that they do not measure the regret that would be felt if system A were implemented but then scenario Y actually materializes. Regret in this eventuality is the difference between the cost that would be incurred by system A under scenario Y and the cost that would occur under scenario Y if system B (the best one under scenario Y) were in place.

Two runs are necessary to answer the robustness question properly in terms of cost regret:

1st Run: Find the best system B under scenario Y; let its total system cost be $TC^*(B|Y)$.

2nd Run: Lock in those aspects of system A's design which could not be changed easily in the event that A were implemented but scenario Y actually occurs; optimize over the remaining aspects of system design using the data of scenario Y. Let $TC^*(A|Y)$ be the resulting total system cost.

$$\text{Regret} = TC^*(A|Y) - TC^*(B|Y), \text{ or}$$

$$\text{Relative Regret} = \text{Regret}/TC^*(B|Y)$$

Notice that some judgement is necessary in the second run. Typically, since facility locations are expensive or organizationally upsetting to change, these locations would be locked up whereas customer zone assignments and transportation flows would not be (they would be optimized during the second run).

The trouble with the first naive approach mentioned is that it compares systems A and B but not their costs. It is quite possible for A and B to look quite similar and yet be far apart cost-wise, or for A and B to look quite dissimilar and yet be very close cost-wise (i.e., for the regret to be small). The trouble with the second naive approach is that it provides no basis for judgement; who can say whether $TC^*(A|Y)$ is "good" unless $TC^*(B|Y)$ is known?

If regret turns out to be small, fine. But what if it turns out to be unacceptable? It will then be necessary to think about how and why systems A and B differ, and to attempt a compromise design which performs acceptably well under both scenarios X and Y. Much depends, of course, on the estimated liklihood that scenario Y will materialize. The concept of regret continues to play a fundamental role.

Example

The model for a major auto parts manufacturer had 9 product groups, 3 plants, 35 candidate distribution center locations, and 144 customer zones. Improved system designs were developed under two major options, namely whether or not there will be a highly automated national "superdepot". Among the robustness studies deemed advisable, it was decided to study the influence of a 25% increase in all transportation rates on Recommended System A (no superdepot).

A first optimization run was done with all transportation rates duly factored by 1.25, and with all distribution centers free to open or close except for the candidate superdepots. The result is schematized below, where "1" stands for an open facility and "0" stands for a closed facility:

Recommended System A	11110001111111110000000000000000000
Optimal System, 1st Run	00001111111111110000000000000000000

Evidently the increased transportation costs induce very substantial changes in the least cost system configuration. It would appear that Recommended System A is not robust in the face of increased transporation costs.

A second run was then done with the same cost structure but with all facility locations fixed per Recommended System A. Thus the only optimization performed in this run was the rebalancing of all transportation flows (especially the facility service areas) in response to the increased transportation costs. The resulting total system cost was, surprisingly, just a shade less than 1% higher than the total system cost under the first run. Thus Recommended System A was shown to be reasonably robust after all in the face of increased transportation costs!

This example illustrates the pitfall of the first naive approach to robustness analysis mentioned earlier. The solution itself changed substantially when transportation rates went up 25%, but in terms of monetary regret this change was relatively insignificant.

D. IMPLEMENTATION PRIORITY ANALYSIS

Among the last runs to be done are those which attempt to provide guidance to management in setting implementation priorities for the various system changes which have emerged as desirable. Toward this end it is appropriate to study the impact of these changes individually and in combination with one another. Generally speaking, those which prove to be responsible for greater improvements in overall system performance should be put into practice before those which yield lesser improvements. Some changes may even prove to be of such small value that their implementation will be postponed indefinitely as not worth the organizational upset, or at least held in abeyance pending restudy with updated data at some future date. Underlying the need for such implementation priority studies are the sobering realities of what is possible in the near term in terms of capital expenditure approval, politics, organizational change and customer acceptance.

It is important to understand that some departures from any given system configuration are usually possible for a negligible cost penalty. For instance, the substitution of Macon for Savannah as the site for a Southeastern distribution center (DC) may well have a very small impact on cost and customer service. The combinatorial richness of DC location patterns, customer zone assignment possibilities, and transportation flows virtually guarantee that the particular system configuration to come out of any given run is just one of numerous alternatives with nearly equal cost. <u>The aim of implementation priority analysis is to exploit this fact</u> by searching among these alternatives for ones that are superior with respect to such hard-to-quantify criteria as those mentioned at the end of the previous paragraph.

These remarks apply with particular force to the important case of evaluating proposed changes in the number and location of DC's. Since the focus is on how the current locational configuration will be changed, it is usually wise to represent the proposed locational configuration as the current configuration plus specific "changes", where each change is one of three types:

(a) an isolated closing of a current DC

(b) an isolated opening of a new DC

(c) the "moving" of a current DC to a new location.

Usually such a representation is evident with the help of a map, although it may not be unique. As a hypothetical example:

Change

1. Close Columbus (current)
2. Close Boston (current)
3. Open Miami (new)
4. Move San Francisco (current) to Los Angeles (new)

The impact of any specific change or combination of changes can be evaluated by doing a run in which the current configuration is modified accordingly. All candidate locations would be locked open or locked closed[5/], but optimization would still take place with respect to customer zone assignments, DC sizes, transportation flows, and plant loadings.

5/ None would be free unless there are discrete size alternatives for a DC at a site that is to be used, and it is desired to allow automatic selection of the best size alternative.

Thus if we compare a base case run locking in the current locational configuration with another run locking in the configuration modified to close Columbus, say, the consequent difference in total system costs is a true and comprehensive evaluation of the cost impact of closing Columbus. It is not just a first approximation of the estimated savings such as one could get by costing out a plausible manual reassignment of Columbus' customers to other nearby DC's; rather, it takes full account of all ripple and interaction effects by making exhaustive use of the adaptive capabilities of the system (i.e., by optimally adjusting customer assignments, DC sizes, transportation flows, and plant loadings). Four runs beyond the base case run will reveal the cost impact of each of the four hypothetical changes taken individually, and a few more runs should suffice to establish the appropriate implementation priority.

Example

The model for a food company had 17 products, 14 plants, 45 candidate distribution center locations, and 121 customer zones (refer to Refs. 1 and 3 for further details). Implementation priority analysis focused mainly on the impact of distribution center location changes on total system cost, since this was the primary management concern and the determinant of capital requirements and organizational dislocations. The most economical reconfiguration required six changes to the current locational configuration. These six changes, which will not be detailed here, were examined individually and in various combinations.

The results of the priority analysis applied to the six changes are summarized in Table 3. The first row indicates that the least cost system has a total normalized cost of 100. Run A shows that the current system at the time of the study had a total cost of 103.15. From run B it can be seen that the total cost associated with the current DC locations could be

Run	DC Locations		Service Areas & Transport.	Total Cost	Differences	
OPT	Optimum (6 changes)		Optimum	100.00		
A	Current		Current	103.15		
B	Current		Optimum	101.43	1.72	save over A
B.1		1	"	101.45	-0.02	
B.2		2	"	101.34	0.09	
B.3	Current & One	3	"	101.14	0.29	save
B.4	Change:	4	"	101.42	0.01	over B
B.5		5	"	101.37	0.06	
B.6		6	"	100.71	0.72	
B.7	Current & Best	1	"	100.01	0.01	loss
B.8	Subset of Changes	2	"	100.12	0.12	over
B.9	Omitting:	4	"	100.13	0.13	OPT
C	Current & Changes 3, 5, 6		Optimum	100.30	1.13	save over B
C.1	Current & Changes	1	"	100.29	0.01	save
C.2	3, 5, 6 and Also:	2	"	100.17	0.13	over
C.3		4	"	100.13	0.17	C
D	Current & Changes 2, 3, 4, 5, 6		Optimum	100.01	0.29	save over C

TABLE 3

Priority Analysis Results for a Food Company reduced by 1.72 by optimizing the DC service areas and transportation flows. Runs B.1-B.6 show the consequence of implementing each of the DC location changes individually. Changes 3 and 6 appear to be quite attractive, changes 2 and 5 only moderately attractive, and changes 1 and 4 are unattractive (presumably because they require the simultaneous presence of another change to be worthwhile). A further consideration was that change 5 would give additional warehouse space in a geographical region where it was particularly needed. Thus changes 3, 5 and 6 were quite attractive at this stage of the analysis.

Another way of looking at the economic value of the more doubtful changes 1, 2 and 4 is to omit (rather than include) them one at a time. This was done in runs B.7-B.9, which selected the best subset of the five changes remaining after each dubious change was omitted. This required placing appropriate candidate locations in a "free" rather than locked status in each run. The results of these runs tend to support the adoption of changes 3, 5 and 6 on a first priority basis. Run C shows, in fact, that changes 3, 5 and 6 together save a little more than would be expected by adding up their one-at-a-time savings ($.29+.06+.72=1.07 < 1.13$).

With changes 3, 5 and 6 decided upon, the other changes were again examined individually. Runs C.1-C.3 reveal that change 1 is still borderline, while 2 and 4 now look worthwhile. This conclusion is further supported by runs B.7-B.9, because it turns out that the results of these runs would have been the same if changes 3, 5 and 6 had been mandatory. In light of this and of other considerations outside the scope of the model, second priority was given to changes 2 and 4 and change 1 was dropped from further serious consideration (run D shows that omitting change 1 incurs an economic penalty of only 1/100 of 1%).

REFERENCES

1. Geoffrion, A.M. and G.W. Graves, "Multicommodity Distribution System Design by Benders Decomposition," *Management Science*, 20, 5 (January 1974).

2. Geoffrion, A.M., "A Guide to Computer-Assisted Methods for Distribution Systems Planning," *Sloan Management Review*, 16, 2 (Winter 1975).

3. Geoffrion, A.M., "Better Distribution Planning with Computer Models," *Harvard Business Review*, 54, 4 (July-August 1976).

4. Geoffrion, A.M., "The Purpose of Mathematical Programming is Insight, Not Numbers," *Interfaces*, 7, 1 (November 1976).

5. Geoffrion, A.M., G.W. Graves, and L. Lee, "Strategic Distribution System Planning: A Status Report," in A. Hax (ed.), *Studies in Operations Management*, North Holland/American Elsevier, forthcoming in 1978.

6. Magee, J.F., *Industrial Logistics*, McGraw-Hill, 1968.

APPENDIX

AN ANALYTICAL MINIMODEL

It is possible under very highly simplified assumptions to write down a "minimodel" that can be solved analytically for just about any result of interest -- the optimal number of warehouses, the annual level of each cost category, the sensitivity with respect to any data coefficient, and so on. The potential advantages of using minimodels in conjunction with full scale mathematical programming models were discussed along with several examples in Ref. 4.

This Appendix presents a minimodel that combines the first two out of the three minimodels given in Ref. 4. We have found it a useful adjunct to a number of different strategic planning models and have therefore selected it as the basis for several results stated in the main text.

The Problem

Consider the case where a single plant with fixed location ships a standard mix of products to a number of warehouses which in turn ship to customers. Make the following assumptions:

(a) Demand is uniformly distributed on the plane with a density of ρ cwt/mi^2 yr.

(b) The inbound freight rate from the plant to each warehouse is r \$/cwt-mi.

(c) The outbound freight rate from each warehouse is t \$/cwt-mi.

(d) The fixed cost for each open warehouse is f \$/yr.

(e) The variable throughput cost of each warehouse is v \$/cwt.

(f) Candidate warehouses are all identical with respect to cost characteristics, have no throughput limits, and are located densely in the plane.

The objective is to select which of the candidate warehouses to open so as to cover a total area of A square miles with non-overlapping equal-sized square service areas at minimal total cost.

The choice of square service areas is a matter of convenience as it can be shown, somewhat surprisingly, that requiring circular or hexagonal service areas would yield the same results for all practical purposes. Assumption (f) and this insensitivity to the exact shape of the service areas imply that, practically speaking, revising assumption (a) to distribute demand uniformly over the continental U.S. rather than over the plane would yield the same results provided at least a dozen or so warehouses are used.

Statement of Main Results

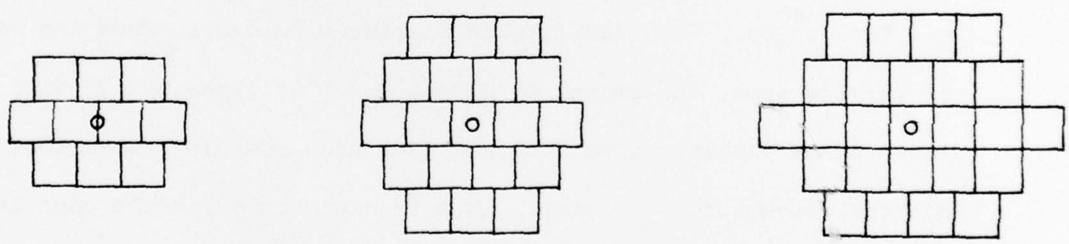
Let n be the number of open warehouses. A good approximation to the value of n which minimizes total annual cost is

$$(1) \qquad n^* = .332A\,(\rho t/f)^{2/3}\sqrt{1-(r/t)^2}$$

so long as $0 \leq r/t \leq 1/2$ and $n^* \geq 10$. The ratio r/t almost always falls toward the low end of this range in practical applications because inbound transportation tends to involve much larger shipments than outbound transportation[1/].

Fractional values of n^* may occur in (1), and to be strictly correct one would then have to compute total cost for the two integers on either side of n^* and choose the better of the two. But for the range of applicability of (1), where $n^* \geq 10$, it seems safe to speak informally of fractional n^*.

The n^* warehouses should be arranged about the plant in such a manner that the union of their (equal-sized) square service areas is, loosely speaking, a tightly-packed figure that resembles as closely as possible a circle centered on the plant with area A square miles. A more precise prescription for optimal arrangement of the n^* warehouses will be given in the course of the proof. Some examples are given in Figure A for selected values of n.



n=10 n=19 n=25

Figure A

Examples of Optimal Service Area Arrangement Relative to the Plant (small circle) for Different Numbers of Warehouses (n)

[1/] For example, a recent issue of *Industry Week* (20 February 1978, p. 80) gives average cent/ton mile figures of 7.0 for truck and 1.4 for rail, which would yield an r/t of 0.2 if all inbound freight goes by rail and outbound by truck. Our practical experience indicates that r/t is usually closer to 0.1 than to 0.2, owing to less-than-truckload shipments and shorter distances traveled on the outbound side.

The warehouses are not located exactly at the center of their respective service areas, but rather are displaced somewhat in the direction of the plant. The optimal amount of displacement is approximately

$$(2) \qquad .572 \left(\frac{r}{t}\right) \sqrt{A/n^*} \quad \text{miles.}$$

Notice that $\sqrt{A/n^*}$ can be interpreted as the length of each side of a service area.

The components of total annual cost associated with the best possible n-warehouse system (n need not equal n*) are as follows:

$$(3a) \qquad nf \qquad \text{fixed warehouse cost}$$

$$(3b) \qquad \rho Av \qquad \text{variable warehouse cost}$$

$$(3c) \qquad \rho Ar(.373\sqrt{A} - .572\,\frac{r}{t}\sqrt{A/n}\,) \qquad \text{inbound cost}$$

$$(3d) \qquad \rho A\sqrt{A/n}\left\{.572\,\frac{r^2}{t} + .3826t\left[1 - \left(\frac{r}{t}\right)^2\right]^{3/4}\right\} \qquad \text{outbound cost.}$$

The corresponding system may not have the optimal number of warehouses, but it is configured optimally relative to the plant given that there must be n facilities. Each facility is displaced optimally from the center of its service area; the amount of displacement is given by (2) with n* replaced by n. Observe that (3c) and (3d) each contain a term that is identical except for the sign. This term accounts for the cost influence of warehouse displacement away from the center of the service areas in the direction of the plant.

From (3) it is straightforward to obtain the associated System Cost vs. Number curve. Leaving aside for the sake of simplicity those costs which do not depend on n (i.e., variable warehouse costs and the first component of inbound cost) and denoting the sum of the remaining costs as

TC*(n), the normalized curve has the following formula:

$$\frac{TC^*(n)}{TC^*(n^*)} = \frac{1}{3}\left(\frac{n}{n^*}\right) + \frac{2}{3}\left(\frac{n}{n^*}\right)^{-1/2} . \tag{4}$$

This relation is graphed in Figure 4. (It would be easy to develop a normalized curve for which TC*(n) includes <u>all</u> cost components, but then the final result would not be independent of the problem data.)

Proofs

The only real task is to justify (3c) and (3d). Relation (1) is obtained simply by setting the first derivative with respect to n of (3a) + (3b) + (3c) + (3d) equal to zero (total cost as a function of n is convex, as follows from the positivity of the second derivative for all $n \geq 0$). Relation (2) and the optimal arrangement of warehouses about the plant will follow from the analysis leading to (3c) and (3d). Relations (3a) and (3b) are immediate. Relation (4) can be demonstrated as follows:

$$\frac{TC^*(n)}{TC^*(n^*)} = \frac{TC^*(n^* \,.\, n/n^*)/n^*}{TC^*(n^*)/n^*}$$

$$= \frac{(n/n^*)f + .3826\rho A^{3/2}\, t\left[1 - \left(\frac{r}{t}\right)^2\right]^{3/4} (n/n^*)^{-1/2}(n^*)^{-3/2}}{f + .3826\rho A^{3/2}\, t\left[1 - \left(\frac{r}{t}\right)^2\right]^{3/2} (n^*)^{-3/2}}$$

$$= \frac{(n/n^*)f + 2(n/n^*)^{-1/2}f}{f + 2f}$$

$$= \frac{1}{3}\left(\frac{n}{n^*}\right) + \frac{2}{3}\left(\frac{n}{n^*}\right)^{-1/2} ,$$

where (1) was used to eliminate the $(n^*)^{-3/2}$ factors in order to obtain the third equality.

An essential step toward justifying (3c) and (3d) is to be able to determine the optimal location of a single warehouse when the locations of its service area and the plant are both given. The situation is depicted in Figure B, where the plant is P miles from the center of an A/n square mile service area and the warehouse is displaced D miles in the direction of the plant.

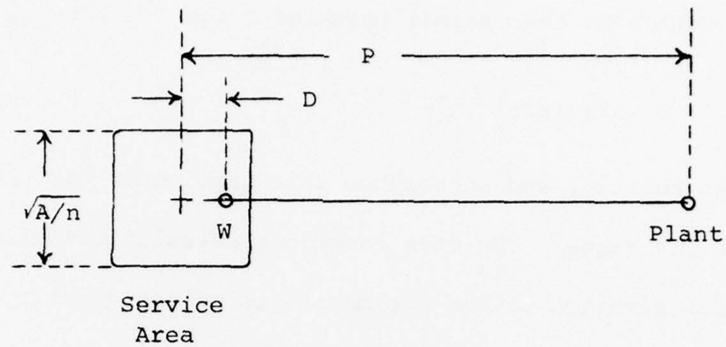


Figure B
A Plant, A Warehouse and It's Service Area

The inbound freight cost for the situation of Figure B is

(5) $$\rho \frac{A}{n} \quad (P - D)\ r$$

and the outbound freight cost is

(6) $$\rho \frac{A}{n} \ E(D\ ;\ \sqrt{A/n})t\ ,$$

where $E(D;\ \sqrt{A/n})$ is the average distance between the warehouse and the customers in its service area (remember that demand is uniformly distributed). An analytical expression for the function $E(\cdot\ ;\ \cdot)$ is known[2/]. This makes it possible to deal with the problem of finding that value of D which minimizes the sum of (5) and (6). The problem can be written after elementary manipulation as

(7) $$\rho\frac{A}{n}\ rP + \rho\frac{A}{n}\ t\ \sqrt{A/n}\quad \underset{D \geq 0}{\text{Min}} \left\{ E\left(\frac{D}{\sqrt{A/n}}\ ;\ 1\right) - \frac{r}{t}\ \frac{D}{\sqrt{A/n}} \right\}\ .$$

[2/] See expression (8.9) on p. 156 of S. Eilon, C.D.T. Watson-Gandy, and N. Christofides, Distribution Management, Hafner Publishing Company, New York, 1971. A partial tabulation appears on p. 157 .

Clearly, the optimal solution depends only on the quantities r/t and $\sqrt{A/n}$, and the second of these can be absorbed into the variable of minimization. Straightforward numerical calculations yield this simple but accurate approximation to the optimal value of D for $0 \leq r/t \leq 1/2$:

$$\text{(8)} \qquad D^* \cong .572\ (r/t)\ \sqrt{A/n}\ .$$

The error is less than 2%, and approaches this level only for r/t toward the upper end of its range. The same numerical calculations also yield the following approximation to the minimal value of the quantity in braces in (7):

$$\text{(9)} \qquad .3826 \left[1 - (r/t)^2\right]^{3/4}\ .$$

The error is less than 1% for $0 \leq r/t \leq 1/2$.

Relations (8) and (9) completely solve the problem of finding the displacement D in Figure B which minimizes inbound plus outbound freight costs over the practical range $0 \leq r/t \leq 1/2$: the optimal displacement is given by (8), the associated inbound freight cost is

$$\text{(5*)} \qquad \rho \frac{A}{n}\ r\left(P - .572 \frac{r}{t}\ \sqrt{A/n}\ \right),$$

and the associated outbound freight cost is

$$\text{(6*)} \qquad \rho\left(\frac{A}{n}\right)^{3/2} \left(.572 \frac{r^2}{t} \quad + \quad .3826t \left[1 - (r/t)^2\right]^{3/4} \right).$$

This proves (2). It also proves (3d), which is just n times (6*). Similarly, the second part of (5*) proves the second part of (3c).

It remains to demonstrate the first part of (3c), namely

$$\text{(10)} \qquad \rho\ A\ r\ (.373\ \sqrt{A}\ )$$

and also to show how the warehouse service areas should be arranged with respect to the plant. These two issues are, in fact, inextricably interrelated. Notice that P depends on the placement of the service area with respect to the plant. Let P^*_{jn} denote this distance for the jth service area under the assumption that n service areas are arranged so as to minimize the sum of their distances from the plant. Then the first part of (5*) yields

$$\rho \frac{A}{n} r \sum_{j=1}^{n} P^*_{jn} , \tag{11}$$

which differs from (10) and hence requires reconciliation. That is, we would like to show that

$$\overline{P}^*(n,A) \stackrel{\Delta}{=} \frac{1}{n} \sum_{j=1}^{n} P^*_{jn} \text{ approximately equals } .373 \sqrt{A} . \tag{12}$$

The quantity $\overline{P}^*(n,A)$ is interpreted as the minimum average distance in miles between the plant and the centers of n square service areas covering a total of A square miles, over all possible non-overlapping arrangements of the service areas.

It is necessary to study how $\overline{P}^*$ varies with n and A. Behavior with respect to A is simple: from elementary dimensional considerations,

$$\overline{P}^*(n,A) = \overline{P}^*(n,1) \sqrt{A} \tag{13}$$

(just exploit the arbitrariness of the mile as a unit of length). Behavior with respect to n is more complicated. Figure A shows service area arrangements that appear to be optimal for n equal to 10, 19 and 25. It is straightforward to verify that $\overline{P}^*(n,A)$ equals $.3686 \sqrt{A}$, $.3698 \sqrt{A}$ and $.3748 \sqrt{A}$, respectively, for these values of n. The geometry of possible arrangements of n service areas is such that $\overline{P}^*(n,1)$ does not behave smoothly as a function of n, but there is considerable regularity and convergence does occur to the asymptotic value $(2/3)/\sqrt{\pi} = .37613$ as $n \to \infty$. The reason for this particular value is that it equals the average distance from the center of a disk of unit area to a uniform distribution of points on it (as n becomes large, the centers of the optimally placed service areas approach a uniform circular distribution for the obvious reason).

Extensive hand calculations show that the convergence of $\overline{P}^*(n,1)$ to the asymptotic value is quite rapid, following the approximate relationship

$$(14) \qquad \overline{P}^*(n,1) \cong .376 \ (1 - 1/7.25\, n) \quad \text{for } n \geq 10 .$$

This estimate is good to within 1% or so. Evidently, there is a "small n effect" which works to diminish inbound freight costs below their asymptotic level, this effect diminishing rapidly as n increases (e.g., the cost (11) is 2% below its asymptotic value for n = 10, 1% below for n = 20, .5% below for n = 40).

The desired relation (12) is now at hand from (13) and (14) if we take n = 18 in (14) as a plausible working approximation. It can be shown that this working approximation introduces a virtually negligible error in the various results which it affects. This completes the verification of relations (1) - (4) and the rationalization of service area arrangement relative to the plant.